Ontology based data warehouse modeling and mining of earthquake data: prediction analysis along Eurasian-Australian continental plates

Shastri L Nimmagadda and Heinz Dreher

Abstract— **Seismological observatories archive volumes of heterogeneous types of earthquake data. These organizations, by virtue of their geographic operations, handle complicated hierarchical data structures. In order to effectively and efficiently perform seismological observatories business activities, the flow of data and information must be consistent and information is shared among its units, situated at different geographic locations. In order to improve information sharing among observatories, heterogeneous nature of earthquake data from various sources are intelligently integrated. Data warehouse is a solution, in which, earthquake data entities are modeled using ontology-base multidimensional representation. These data are structured and stored in multi-dimensions in a warehousing environment to minimize the complexity of heterogeneous data. Authors are of the view that data integration process adds value to knowledge building and information sharing among different observatories. Authors suggest that warehoused data modeling facilitates earthquake prediction analysis more effectively.**

I. Introduction

The Earth's surface is broken into thin jigsaw-like pieces called tectonic plates. The Australian continent is actually part of the Indian-Australian plate as described in [1], [4] and [9]. Figs. 1 and 3 show the seafloor and land shape (topography) of the Eurasian - Australasian regions. Earthquakes are most common at plate boundaries, where tectonic plates move and meet.

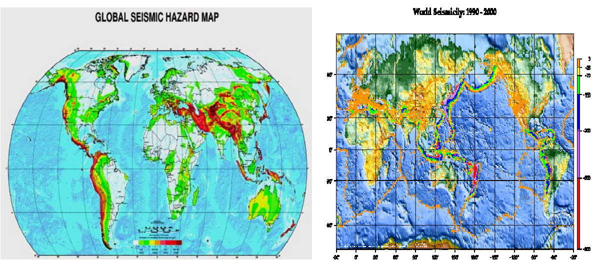

Fig.1. Seismically active and hazard provinces – distribution of continental plates

80% of all recorded earthquakes occur around the edge of the Pacific Plate, affecting the Pacific countries (Fig.3). Earthquake vibrations travel very fast; up to 14 km/sec.

Shastri L Nimmagadda is with Wafra Joint Operations, Kuwait Gulf Oil Company, Kuwait City (e-mail: shastri@eftel.net.au). Heinz Dreher is with School of Information Systems, Curtin Business School Curtin University of Technology, Perth, WA, Australia (e-mail: h.dreher@curtin.edu.au).

The science of earthquakes, seismology, as narrated in [6] and developed in 1880 (John Milne invented the first proper seismograph) enables accurate recording (Fig. 2) of earthquakes.

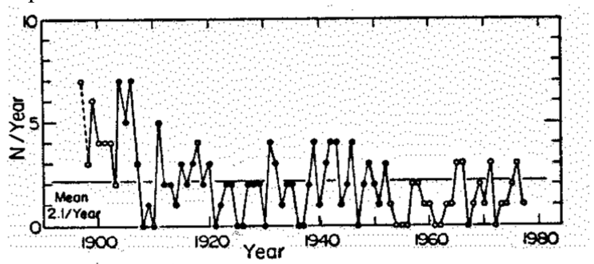

Fig.2. Earthquake occurrences last century

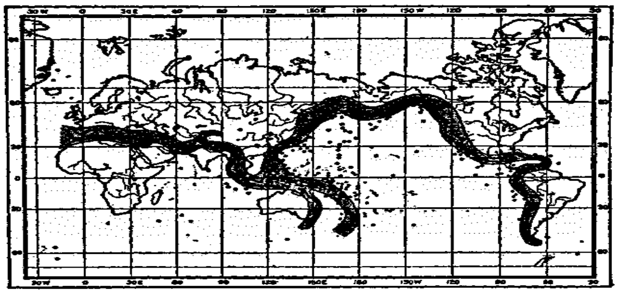

Fig.3. Seismic map showing the two earthquake belts of the world [4]

Analyses of earthquake data ([6] and [9]) reveal that around 20 earthquake-events >= Richter 7.0 occur annually, and whilst this has remained constant the *magnitude* of quakes from 1900-1920 had about *twice as great* an annual energy release as the whole of period up to 1977 (Fig.4). Volumes of these earthquake data are recorded and documented in bits and pieces over the last few centuries with a hope that an analysis may lead to insight and, prediction, and aversion.

II. Problem Definition

Two major issues are adversely affecting the prediction of earthquakes; (1) effective access to distributed unstructured earthquake information (both Web and offline sources); and (2) lack of infrastructure for precise and accurate earthquake information access and retrieval. If the earthquake data are not appropriately integrated, information that is shared by various seismological observatories may lead to failure of knowledge building and prediction.

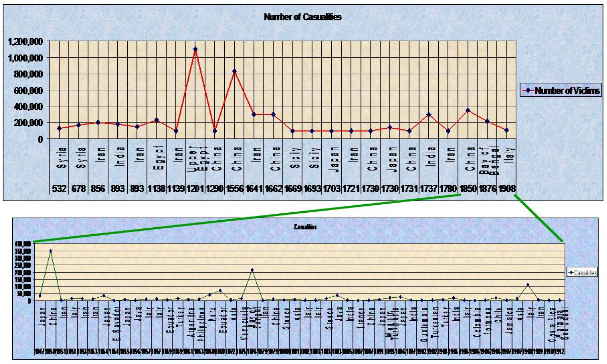

Fig.4. Comparison among 15 centuries earthquake casualty data

We attempt to use a systematic shared ontology, supporting data warehouse modelling and data mining research to organize and store valuable earthquake data. Most of the published results are available on the Web, through journal databases and earthquake databases in several formats and on software platforms, and are thus amenable to this approach.

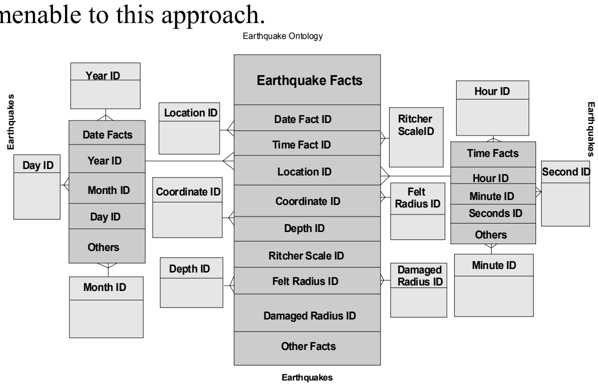

Fig.5. Ontology-base star type multidimensional model for warehousing earthquake data

Understanding an earthquake system ([1], [4], [6] and [9]) and its prediction is a significant problem. Data integration and information sharing among different earthquake systems of different continental plates are key issues of the present problem definition. Little attention has been paid in integrating and organizing these historical earthquake data. Ontology (Fig. 5) for structuring data warehouse and techniques for mining the warehoused data are needed for solving problems, associated with this type of data integration and analysis. To date, there has been no systematic investigation of these earthquake data volumes using the proposed, focused technologies. Meticulous analysis of earthquake systems of different continental plates at different periods and geographic locations is much-needed research. Most of the earthquake data are unorganized and in fact, massive stores of these data hide with undiscovered (unknown knowledge or intelligence) data patterns. Interpreting the patterns, correlations and trends among earthquake data and finding a possible predictive solution, is the goal of our current research. Through accurate prediction, the enormous loss of life and property can be averted.

III. Issues of Earthquake Prediction

Authors experience the following major issues affecting earthquake predictions:

- Handling numerous entities and attributes, mapping and modelling hundreds of tables is a tedious process;
- Data integration (sometimes among 10-15 seismological observatories) of enormous amount of multi-disciplinary data. Communicating and integrating data among multiple observatories must be a prerequisite for the successful design, development and implementation of prediction systems;
- Management and maintenance of large earthquake databases are big issues. Periodically, earthquake data volumes of different continental plates are accumulated in unmanageable form, especially in the knowledge domain.

Earthquake data mining and knowledge building for prediction purposes:

- Knowledge-building from these massive data structures is an intricate issue. With increasing volumes of periodic data, there is difficulty in understanding or retrieving knowledge without reference to an overarching earthquake system;
- Inconsistency among earthquake metadata events and characteristics affect the earthquake prediction analysis. Unknown and unpredictable earthquake data patterns may affect extraction of predictive knowledge. Improved earthquake data management may reduce ambiguity in prediction analysis;
- Applicability and feasibility of data warehouse supported by ontological modelling, and a combined application of data mining and visualization can have a tremendous impact on earthquake system knowledge discovery and thus facilitate prediction analysis.

Ontology, as described in [2], [3], [5] and [7], provides basic knowledge and semantic information for designing a data warehouse. Mining of data views and interpreting them in knowledge domains are key areas of current research. Database designers may use conceptual modeling to design hundreds of fact and dimension tables using relational and hierarchical structures. Unless the true meaning of the entities and their relationships are understood, the conceptual model representing the earthquake data and the business rules applied, while mapping the entities and relationships, is incomplete. Inconsistencies or errors that occur during database design are expensive to fix at later stages of warehouse development. We propose ontology to help ensure consistency in the semantics and models built based on multiple dimensions and entities. Data warehouse maintenance and administration are other issues that need to be examined and judged while developing an ontology or conceptual model. Ontology based warehouse approach plays a key role in gaining a holistic understanding of earthquake data. Models based on relationships and associations or properties of earthquake data entities (Figs. 5 and 6) and dimensions affect characterization of the seismological data integration process and thus prediction. This may also undermine the data warehouse maintainer's ability to understand the complexity of hierarchical, relational or network data structures.

End users of a data warehouse may also employ conceptual models to help formulate the queries or updates of databases. If semantics of earthquake data and the relationships among their entities are not understood, users may not have realized the significance of ontology in a data warehouse design. Ontology and semantics of earthquake data are very complex (especially on period and location dimension hierarchies). Key entities are considered for understanding the relationships among several seismological observatory entities (Figs. 4-6). Based on the observatory activities, the hierarchical structure view with root *earthquake*, a generalized super-type entity as a subject-oriented and, from which other specialized associated entities such as *location, position, period* and *precise time of occurrence* sub-types are derived, while building knowledge-base data relationships. The present study embodies the scenarios of application of ontology and achieving its purpose of earthquake data integration for a possible prediction system.

A. Warehouse Modeling of Earthquake Ontology

The following steps are needed:

1. Acquire data that attribute to earthquake system;
2. Identify entities and attributes of earthquake data;
3. Build relationships among entities with their common attributes;
4. Structure and de-structure complex relationships among data entities (dimensions, in multidimensional structuring);
5. Acquire *locations, periods, magnitudes* (Richter scales), *damaged radius* and *felt radius* data;
6. Represent all the entities into relational, hierarchical and networked data structures;
7. Develop conceptual models using ontology;
8. Integrate earthquake data using ontology-base multidimensional warehouse modeling for prediction knowledge.

As demonstrated in [7] and [8], ontology application in petroleum industry evaluates the inventory of the petroleum reserves in the Western Australian production basins. Hundreds of entities and their attributes associated with petroleum systems and their elements are well organized ontologically and semantically in a warehousing approach, so that basin and petroleum system knowledge is well understood for managing petroleum reserve inventories. Similar approach is used to integrate the current earthquake data.

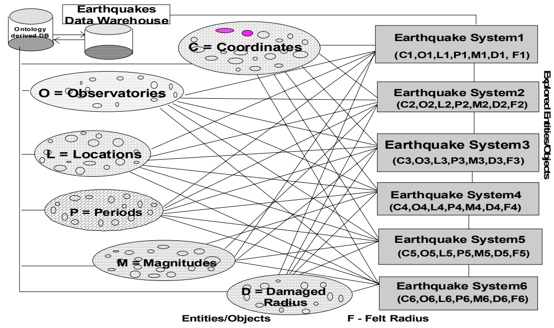

Fig.6. Earthquake Systems Analysis

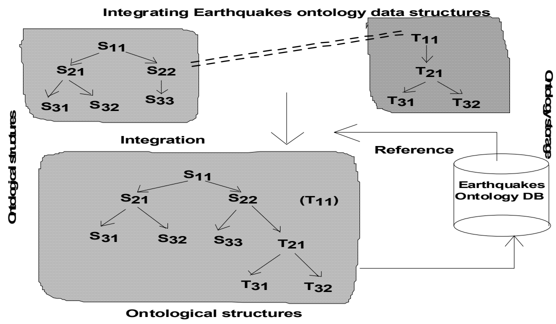

Fig.7. Integration of Ontology data structuring

Integration of all entities and their relationships is a part of ontology design and development for an effective data warehouse modeling and mining of the earthquake data as narrated in Fig. 7. An integrated warehouse modeling frame-work as shown in Fig. 8, integrates data from several dimensions (in knowledge domain). Warehouse frameworks expect these data, stored in several dimensions. These dimensions and their instances are connected and integrated from several earthquake systems and their recording systems. Earthquakes are recorded through multiple seismographs. A seismograph has four basic components; a seismometer to detect the ground motion, an amplifier to boost the signal, a clock (which measures as accurate up to 0.01 seconds) used to time the arriving seismic waves for later analysis and a recording device. All these components are stored through multi-dimensions representation in a data warehouse.

As described earlier, the Asian-Australian plate is active and because of its deep seated earthquake occurrence and instances, since recent past, this region has become naturally disastrous areas of destruction on this planet – the most recent very significant event being the 26th December 2004 tsunamis triggered by the Sumatra-Andaman earthquake centered off the west coast of Sumatra, Indonesia.

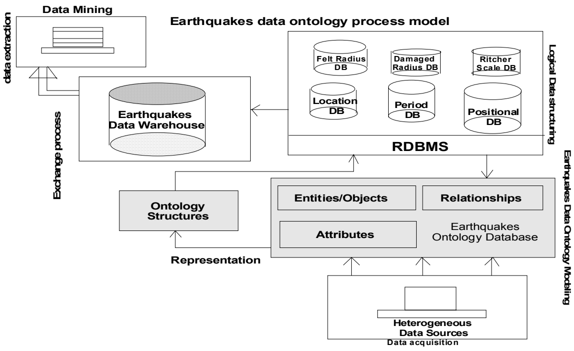

Fig.8. Warehouse-base Ontology framework

Indonesia, Australia, Thailand, Malaysia, India, Pakistan and China countries fall along this Eurasian-Indian-Australian plate, which experience major deep seated earthquakes (Fig.9) and document several records. Indian-Australian plate is being pushed north and is colliding with the Eurasian, Philippine and Pacific plates. This phenomenon conceptually exists in the data. These inherent data are extracted in the data mining process stage.

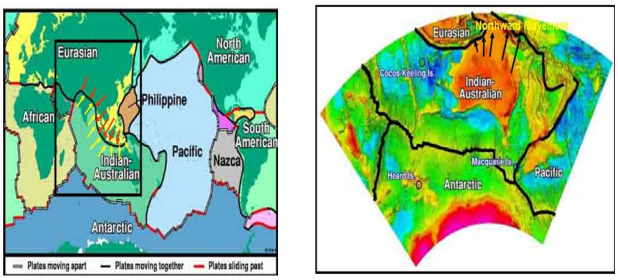

Fig.9. Indian-Australian plate collision with the Eurasian plate; Australian plate moving north

Data mining extracts hidden intelligence from warehoused data. For example, in Australia, Adelaide has the highest earthquake hazard of any capital city, with more medium-sized earthquakes in the past 50 years than any other. South Australia is being slowly squeezed sideways by about 0.1 mm/yr. We can't predict when earthquakes happen but measuring these changes, combined with Adelaide's earthquake history, helps us understand when the next big earthquake might happen. Australia's largest recorded earthquake occurred in 1941 at Meeberrie, WA. Its *magnitude* was estimated to be 7.2 but fortunately it occurred in a remote area. A *magnitude* 6.8 earthquake occurred at Meckering in 1968, causing extensive damage to buildings, and was felt over most of southern WA. Earthquakes of *magnitude* 4 or more are fairly common in Western Australia, with one occurring approximately every five years in the Meckering region – for earthquakes in the Perth region consult [9]. These hidden data in knowledge domain are interpreted for possible earthquake prediction analysis.

IV. Analysis and Discussions

We continue to be asked by many people throughout the world whether earthquakes are on the increase. Although it may seem that we are having more earthquakes, those of *magnitude* 7.0 or greater have remained fairly constant. A partial explanation may lie in the fact that in the last twenty years, we have definitely had an increase in the number of earthquakes; we have been able to locate them each year. This is because of an increase in the number of seismograph stations in the world, accuracy of measurements and many improvements in global communications. In 1931, there were about 350 stations operating in the world; today, there are more than 8,000 stations and these data are recorded and documented rapidly from these stations by electronic mail, internet and satellite means. This increase in the number of stations and timely receipt of data, have allowed seismological observatories to locate and record earthquakes more rapidly and also detect many smaller earthquakes which were undetected in earlier years. The NEIC (USA) now locates about 20,000 earthquakes each year or approximately 50 per day. According to long-term records (since about 1900), we expect about 17 major earthquakes (7.0 - 7.9) and one great earthquake (8.0 or above) in any given year.

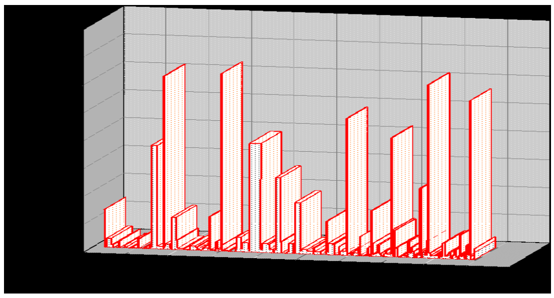
Fig.10. Cyclic nature of earthquakes (?)

In the above graph, *period* and *depth* of occurrence are plotted in a 3D plot view. Fig. 10 shows two key significant findings, analyzing past 5 years historical data, occurrence of earthquakes appears to be cyclic and past couple of years, occurrence of earthquakes is strong.

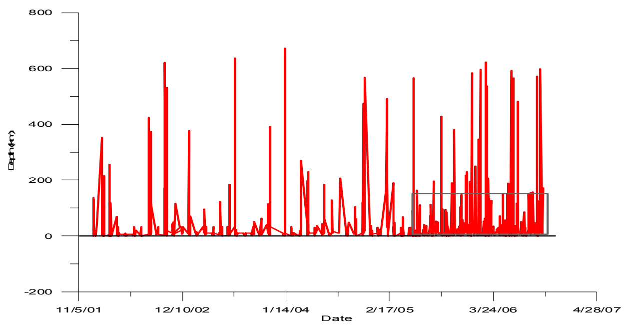

Fig.11. 2D view between period and depth dimensions

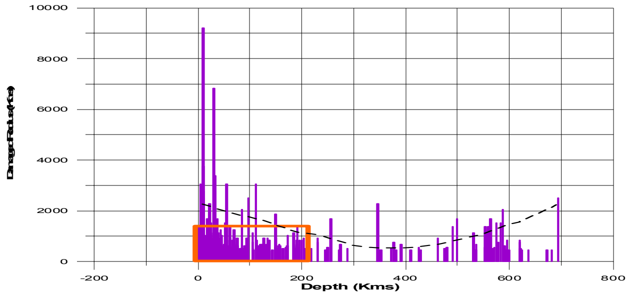

Fig.12. 2D view between depth and damaged radius dimensions

As shown in Figs. 11 and 12, *period* and *depth* dimensions are plotted. Again, occurrence of earthquake is interpreted to be cyclic and their occurrence is strong, more in recent years. A plot drawn between *depth* and *damaged radius*, indicates that at shallower depths there is more damage in square kilometer area.

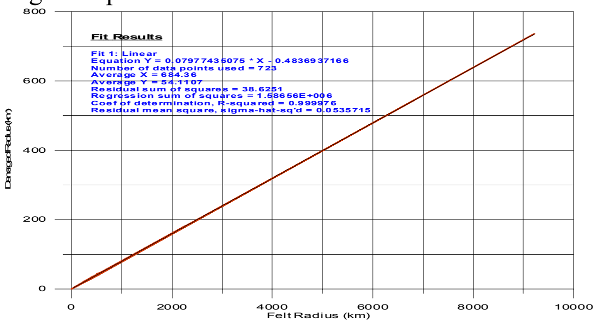

Fig.13. A linear relationship between felt radius and damaged radius dimensions

As shown in Fig. 13, there is a linear relationship between *felt radius* and *damaged radius* dimensions, as anticipated, with more sense of earthquake, there is corresponding response of damaged effect. Plot drawn between *magnitude* of earthquake and *damaged radius* dimensions, shows that there is an exponential relationship (Figs. 14-15).

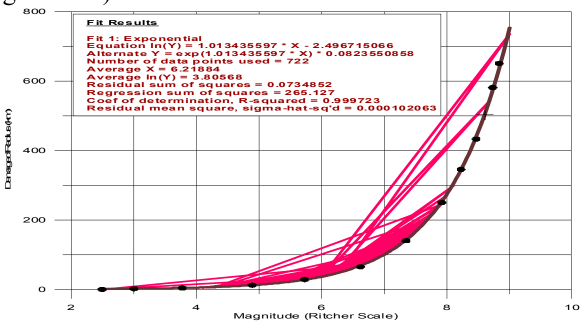

Fig.14. Exponential relationship between magnitude and damaged radius dimensions

It is interesting to observe that at smaller *magnitudes, damaged radius* response is gentler up to depths of 300kms, compared to higher magnitudes, which have sharp and steeper damaged responses up to depths of 700kms.

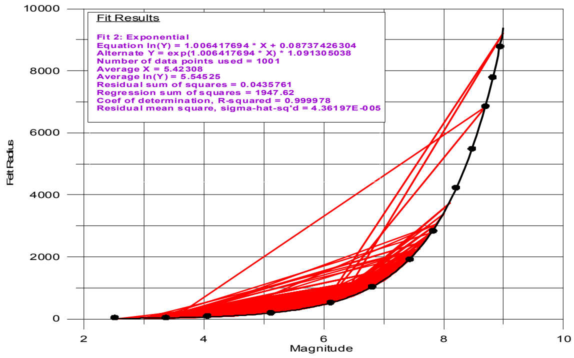

Fig.15. Exponential relationship between magnitude and felt radius dimensions

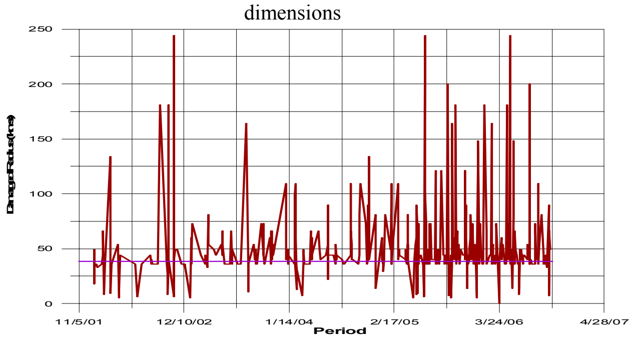

Fig.16. 2D view plot between period and damaged radius dimensions

Most of the responses (Figs. 11-12) of *felt radius* are stronger up to depths of 300kms with 3-8 *magnitude* ranges. Earthquakes, which occurred during 2003 are stronger compared to those of 2004. Again *damaged radius* response is effective during 2005-06. A similar phenomenon is observed in the plot drawn between *period* and *magnitude* dimensions. World is seismically very active in the recent years, especially due to the Indian-Australian and Eurasian continental plates collision. As shown in Figs. 16-17, with increase in *periods* and the *number* of earthquakes (dimensions), increase in *magnitude* and *frequency* are interpreted, which could be due to the occurrence of more earthquakes during the Indian-Australian and Eurasian plates' collision.

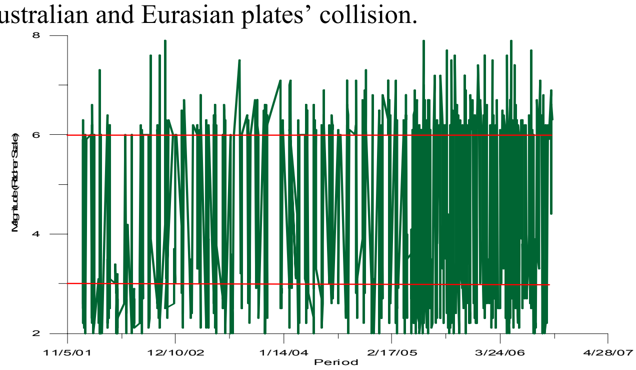

Fig.17. A data view of warehoused earthquakes volume – showing intensity and frequency of earthquake magnitudes during period 2001-06

Figs. 18-19 show *magnitude* of earthquakes of the order of 6, which is almost constant at different depths of occurrence in between 0-300kms. The rate of change in *magnitudes* is very high at shallower depth of 50-100kms.

Bubble clusters and their sizes vary as per scale of earthquake *magnitude*, damaged and felt radii and depth attributes instances. These graphs are representation of 3 dimensional attribute instances and the 3rd dimension lies with the bubble size. Bubble clusters and their magnitudes are changing, depending on the multidimensional attribute instances.

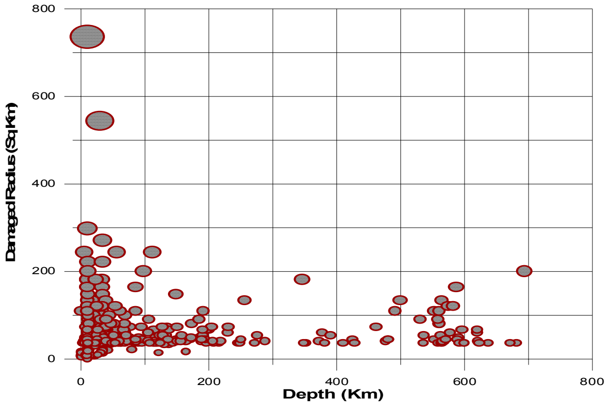

Fig.18. Multidimensional bubble plot view among depth, damaged and felt radii dimensions

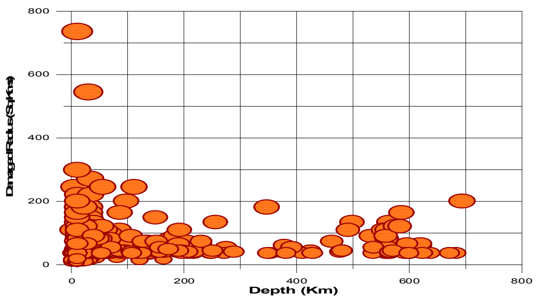

Fig.19. Multidimensional bubble plot view among depth, magnitude and damaged radius dimensions

V. Conclusions and Recommendations

Results and discussions are inconclusive. Using ontology based data warehouse approach, processing and interpretation of earthquakes associated with Indian subcontinent and its region, such as Pakistan, China, Indonesia and Philippines are under study. More rigorous approach is recommended to make a definite inference and conclusion on prediction analysis. Multidimensional modeling approach of warehousing the earthquake data facilitates an effective mining of heterogeneous data more robustly at different varied dimensions. The authors are of the view that a systematic earthquake analysis will facilitate the augmentation of the prediction analysis. It is too early to conclude the prediction calculations, because of poor organization of earthquakes and their understanding. More data are being acquired for building sensible and more interpretable models that narrate more meaningful and reliable prediction knowledge.

Mass population may never experience small-scale earthquakes, in spite of common occurrence of hundreds of minor earthquakes on this planet. So today -- somewhere -- an earthquake may have occurred. It may be so light that only sensitive instruments will perceive its motion; it may shake houses, rattle windows, and displace small objects; or it may be sufficiently strong to cause property damage, death, and injury. It is estimated that about 700 shocks each year have this capability when centered in a populated area. The problem, however, is in pinpointing the area where a strong shock will centre and when it will occur. But in technology world, this notion is changed and all these natural disasters can be predicted, if the earthquakes are systematically analyzed with ontology-base warehouse modeling. Just as the weather bureaus now predict hurricanes, tornadoes, and other severe storms, seismological observatories and other geophysical institutes can predict earthquakes by their systematic analysis and evaluation. Ontology based warehouse modeling facilitates data mining more efficiently ([7] and [8]) and authors deduce that similar earthquake data modeling approach will assist the earthquake prediction analysis more effectively.

References

[1] Abe, K. and Kanamori, H., (1980) Historical Earthquakes, *Tectonophysics*, Vol. 62, p. 196.

[2] Anahory, S. and Murray, D. (1997) *Data warehousing in the real world: a practical guide for building decision support systems*, Pearson Education, pp. 255-360.

[3] Berson, A. and Smith, J. S (2004) *Data warehousing, data mining & OLAP*, Mc Graw – Hill Edition, pp. 205-219, 221-513.

[4] Daly, R.A., (1926) *Our Mobile Earth*, New York & London, p. 6

[5] Hoffman, D.R (2003) *Effective Database Design for Geoscience Professionals*, PennWell Publishers, pp. 205-222.

[6] Markus Båth, *Introduction to Seismology*, 2nd ed., Basel, Boston, Stuttgart 1979, pp. 27, 262-266.

[7] Nimmagadda, S. L and Dreher, H. (2006) Ontology-Based Data Warehousing and Mining Approaches in Petroleum Industries: in Nigro, H.O., Císaro, S.G., and Xodo, D. (Eds.), *Data Mining with Ontologies: Implementations, Findings and Frameworks*, a book to be published in 2007 by Idea Group Inc. http://www.exa.unicen.edu.ar/dmontolo/

[8] Nimmagadda, S.L. and Dreher, H. (2006a) Mapping and Modeling of Oil and Gas Relational Data Objects for Warehouse Development and Efficient Data Mining, a paper presented and published in the *Proceedings of 4th International Conference in IEEE Industry Informatics*, held in Singapore, August

[9] Trevor J. Middelmann, M. and Corby, N. (2005), Cities Project Perth Report, in Sinadinovski et al. *Chapter – 5, Earthquakes Risk, GA Publications – Earthquakes Risk Assessment and Management in Australia*